Assignment II

February 27, 2025

ID5002W: Industrial AI Laboratory

Total Marks: 80

Instructions

1. Assignment shall be submitted before the due date. Late submissions will not be entertained. If you cannot submit the assignment due to some reasons, please contact Dr. Tirthankar by email.

2. All the assignments must be the student's own work. The students are encouraged to discuss or consult friends or classmates. However, they have to submit their own work. Any malpractice will be reported to the authorities and action will be taken as per the IIT Madras rules.

3. If you find the solution in the book or article or on the website, please indicate the reference in the solution

4. You are expected to submit "*.py" file instead of notebook. Follow the steps mentioned below:

   (a) Use the datasets provided to you as per your roll number for solving the assignment problems on regression and classification.

   (b) Create one notebook for Problem 1 and another notebook for Problem 2.

   (c) If you have used

   ```
   df = pd.read_csv(roll_no.csv)
   ```

   where "roll_no.csv" is the file provided to you, replace the code with the following before submitting the assignment.
   For Problem 1,

   ```
   dataset = "./data/classification/roll_no.csv"
   df = pd.read_csv(dataset)
   ```

   For Problem 2,

   ```
   dataset = "./data/regression/roll_no.csv"
   df = pd.read_csv(dataset)
   ```

   (d) Now, download each notebook as a "*.py" file and submit both the "*.py" files.

5. A short report in pdf format along with the code should be submitted containing results and analysis as asked in the assignment.

6. Please note:

   - Code should execute without any error.
   - The code should be clean with readable comments.
   - The output of the code should be in a relevant format so that it can be understood by an evaluator.

7. Grading Policy:

   - 50% code correctness.
   - 10% code readability and comments.
   - 40% on report and analysis.

Problem 1

Please use the dataset provided for this problem (regression) and polynomial features of degree 3:

(a) Implement a Ridge regression model (Default hyperparameters) (**Marks: 4**)

(b) Implement a Lasso regression model (Default hyperparameters) (**Marks: 4**)

(c) Report the impact of the regularization term (α) on the coefficients' values and the model's performance. Compare both models and elaborate on your findings. (**Marks: 12**)

(d) Create polynomial regression models of increasing degree (3-10), calculate and compare their MSEs, and discuss bias, variance, overfitting, and underfitting. (**Marks: 12**)

(e) Create a kernel ridge regression model with $\alpha = 1$ and evaluate the impact of different kernels. (**Marks: 8**)

Problem 2

Please use the dataset provided for this problem (classification). The feature columns are $x_0, x_1, x_2, \ldots, x_{26}$ and the class column is y.

(a) Classify the dataset using Logistic regression and find summary statistics for important features. (**Marks: 8**)

(b) Also implement the LDA and QDA and evaluate the model performance. (**Marks: 8**)

(c) Implement stochastic gradient descent classifier from scratch (use appropriate loss term for classification). Also report the best learning rate based on a hyperparameter search. (**Marks: 24**)